# RESEARCH REPORT

ISSN 1172-8531



# CAUSAL DIAGRAMS FOR $I(1)$ STRUCTURAL VAR MODELS

Granville Tunnicliffe Wilson† and Marco Reale*

* *Department of Mathematics & Statistics,*
*University of Canterbury,*
*Private Bag 4800, Christchurch, New Zealand.*

**Report Number**: UCDMS2002/6 July 2002



Department of Mathematics

UNIVERSITY OF

# CAUSAL DIAGRAMS FOR $I(1)$ STRUCTURAL VAR MODELS

**Granville Tunnicliffe Wilson**[†] **and Marco Reale**[*]

[*] *Department of Mathematics & Statistics,*
*University of Canterbury,*
*Private Bag 4800, Christchurch, New Zealand.*

**Report Number**: UCDMS2002/6 July 2002

**Keywords:** Causality, directed graphs, conditional independence, multivariate time series, structural vector autoregression

[†] *Department of Mathematics & Statistics, Lancaster University, Lancaster LA1 4YF, UK.*

# Causal diagrams for $I(1)$ structural VAR models*

Granville Tunnicliffe Wilson[†] and Marco Reale[‡]

August 1, 2002

**Abstract**

Structural Vector Autoregressions allow dependence among contemporaneous variables. If such models have a recursive structure, the causal relation among the variables can be represented by directed acyclic graphs. The identification of these relationships for stationary series may be enabled by the examination of the conditional independence graph constructed from sample partial autocorrelations of the observed series. In this paper we extend this approach to the case when the series follow an $I(1)$ vector autoregression. We show that, even though the theoretical partial autocorrelations are undefined for integrated processes, exactly the same data procedures and sampling properties may be applied. The theoretical reasoning is supported by the empirical results of simulation, and applications from banking series and term interest rates are used to illustrate the procedure.

*JEL Classification: C10, C32.*
*KEY WORDS: Causality, Directed graphs, Conditional independence, Multivariate time series, Structural vector autoregression.*

*The authors wish to thank Clive Granger and Daniel Peña for their useful comments.

[†]Department of Mathematics and Statistics, Lancaster University, Lancaster LA1 4YF, UK.

[‡]Department of Mathematics and Statistics, University of Canterbury, Private Bag 4800 Christchurch, New Zealand.

# 1 Introduction

## 1.1 Structural vector autoregressions.

In the recent literature on causal analysis there have been many developments based on graphical modeling (e.g., Pearl, 2000 and Spirtes et al., 2000) where causal relations among random variables are described by directed acyclic graphs (Glymour and Spirtes, 1988, pp 179–181). This approach has also been extended to time series analysis both in the frequency (Dahlhaus, 2000) and the time domain (e.g., Swanson and Granger, 1997 and Lauritzen and Richardson, 2002).

In particular in the latter context Reale and Tunnicliffe Wilson (2001, 2002) considered the $p$th order vector autoregressive model VAR($p$) of a stationary $m$-dimensional time series $x_t = (x_{t,1}, x_{t,2}, \ldots, x_{t,m})'$ in its structural form:

$$\Phi_0 x_t = d + \Phi_1 x_{t-1} + \Phi_2 x_{t-2} + \cdots + \Phi_p x_{t-p} + a_t \tag{1}$$

where $d$ allows for a non zero mean of $x_t$ and $a_t$ is a multivariate white noise with variance matrix $D$.

One requirement of this model is that $D$ is diagonal. A further condition, on $\Phi_0$, is that it represent a recursive (causal) dependence of each component of $x_t$ on other contemporaneous components. This is equivalent to the existence of a re-ordering of the elements of $x_t$ such that $\Phi_0$ is upper triangular with unit diagonal. Thus $x_{t,i}$ depends upon all the remaining contemporaneous variables $x_{t,i+1}, \ldots x_{t,m}$. It is convenient for us to take this as our standard ordering, although the order of causation, $x_{t,m} \rightarrow x_{t,m-1} \rightarrow \cdots \rightarrow x_{t,1}$ is the reverse of the natural ordering, We call $x_{t,1}$ the first and $x_{t,m}$ the last variable.

The model may be transformed to its canonical form:

$$x_t = c + \Phi_1^* x_{t-1} + \Phi_2^* x_{t-2} + \cdots + \Phi_p^* x_{t-p} + e_t, \tag{2}$$

which does not allow dependence among contemporaneous variables, by dividing through by $\Phi_0$, in which form $x_t$ is expressed as a linear combination of $x_{t-1}$, $x_{t-2}$, ..., $x_{t-p}$ with an error term $e_t = \Phi_0^{-1} a_t$. This is the linear innovation, having variance matrix $\Sigma$ related to $D$ by

$$\Sigma^{-1} = \Phi_0' D^{-1} \Phi_0. \tag{3}$$

The model (1) is not unique, in that the transformation to the canonical form, which is unique, may be reversed by the choice of *any* matrix $\Phi_0$ which satisfies (3). Each possible ordering of the series gives a different form of (causal) structural model. In Reale and Tunnicliffe-Wilson (2001) we present a methodology that is based on the assumed existence of a model representation (1) that is sparse, i.e. has a relatively small number of coefficients. We will call this the *true* structural model. The method seeks to explore and identify the ordering that corresponds to such a sparse representation. It may reveal more than one such form. The concept of a true model may be idealized, and the method may be viewed as seeking an approximation to the ideal, but it is a useful concept for developing the properties of the procedure. If the number of series is small, each of the different possible causal orderings of contemporaneous dependence could be explored directly, but for a larger number of series, this would be impractical. Our method then offers a practical way forward.

## 1.2 Causal representation.

The model (1) may be represented by a directed acyclic graph (DAG) in which the components of $x_t$, $x_{t-1}$, $\ldots$, $x_{t-p}$ form the nodes, and causal dependence is indicated by arrows linking nodes. The nature of the model is that all arrows end in nodes representing the contemporaneous variables on the left hand side of (1). Some arrows will start from the past, and some from other contemporaneous variables. As an illustration, we reproduce a structural model presented by Reale and Tunnicliffe-Wilson (2001).

The data were 8 years of monthly values of three variables of the Italian monetary system: the re-purchase agreement interest rate, $x_{t,1}$; the average interest rate on government bonds, $x_{t,2}$ and the average interest rate on bank loans, $x_{t,3}$. The DAG representing the structural VAR(2) model, chosen by our procedure, for these series, is shown in Figure 1. The numbers attached to the links are the coefficients in the linear predictor for the corresponding contemporaneous variable.

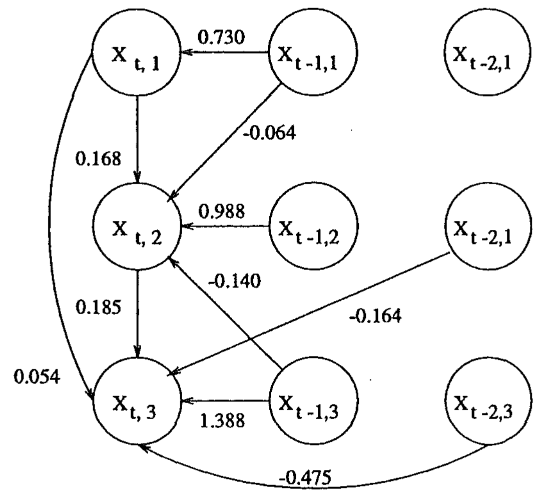


Figure 1: The DAG representation of the structural VAR(2) identified for the Italian monetary variables.

These coefficients are estimated by single equation ordinary least squares (OLS) regression. This is fully efficient under the working assumption, confirmed by a normality test, that the vector series is Gaussian. Our methods are also applicable, and the properties of the estimates given by the regression are reliable, under wider conditions, such as $e_t$ being I.I.D., presented for example in Anderson (1971).

## 1.3 An identification procedure.

The exploratory methods, described in Reale and Tunnicliffe Wilson, 2001, are as follows. The first step is to identify the overall order $p$ of a VAR model for the series. The second and central step is to construct a sample conditional independence graph (CIG) for the variables $x_t$, $x_{t-1}, \ldots, x_{t-p}$ which form the nodes of the graph. Being based upon statistical correlations, the only causality which can be attached to this is that indicated by the arrow of time. The third step is to determine which DAG representations of structural VAR models are consistent with this CIG. Typically, there will only be a small number of such possibilities, if the CIG is sparse, i.e. contains a relatively small number of links. In particular, the graph may admit only a small number of possible interpretations of the direction of dependence between contemporaneous variables. The final step is to fit the corresponding structural VAR models by regression and select one using a criterion such as AIC (Akaike, 1973).

These exploratory procedures rely upon the assumption of stationarity of the vector time series that is being modeled, because they are expressed in terms of sample values of partial autocorrelations of the series. The partial autocorrelations of a time series, and the autocorrelations from which these are derived, are only defined for a second order stationary process. But we wish to apply the same procedures to time series that appear to be $I(1)$ processes with an arbitrary degree of cointegration.

For example Figure 2 shows three US dollar term interest rates over a period of 600 days, that are similar to random walks with the appearance of cointegration.

In Tunnicliffe Wilson, Reale, and Morton, (2001), our methods for structural autoregressive modeling are extended to the construction of a structural ARMA(1,1) model for seven US dollar term interest, including those illustrated in Figure 2. In doing this it was assumed that the term rates were $I(0)$. This is a reasonable assumption for very long records of the series, because they can be assumed to be bounded. One would however, find it hard to reject the hypothesis that series such as these followed an $I(1)$ model, even for the relatively long records illustrated. We will describe such series as *near*-$I(1)$.

Distributional properties of estimates which may in theory be valid asymptotically for any stationary process, may in practice be far from the asymptotic properties for quite long samples of *near*-$I(1)$ processes. An example arises from the univariate AR(1) process. The standard $t$ value for the autoregressive coefficient $\phi$ has an asymptotic "standard" Normal$(0, 1)$ distribution under wide assumptions. However, if $\phi$ is close to 1, in quite large samples the distribution can be closer to that of the "non-standard" Dickey Fuller distribution that arises in the $I(1)$ case for $\phi = 1$ (Abadir, 1995).

## 1.4 Extension to $I(1)$ processes.

The main point of this paper is to show that the statistical procedures that we presented in Reale and Tunnicliffe Wilson, 2001, can be applied, without change, to vector $I(1)$ processes. The results that we shall present, re-assure us that we can base our exploratory inference for $I(1)$ and *near*-$I(1)$ processes, on standard normal distributions, and that

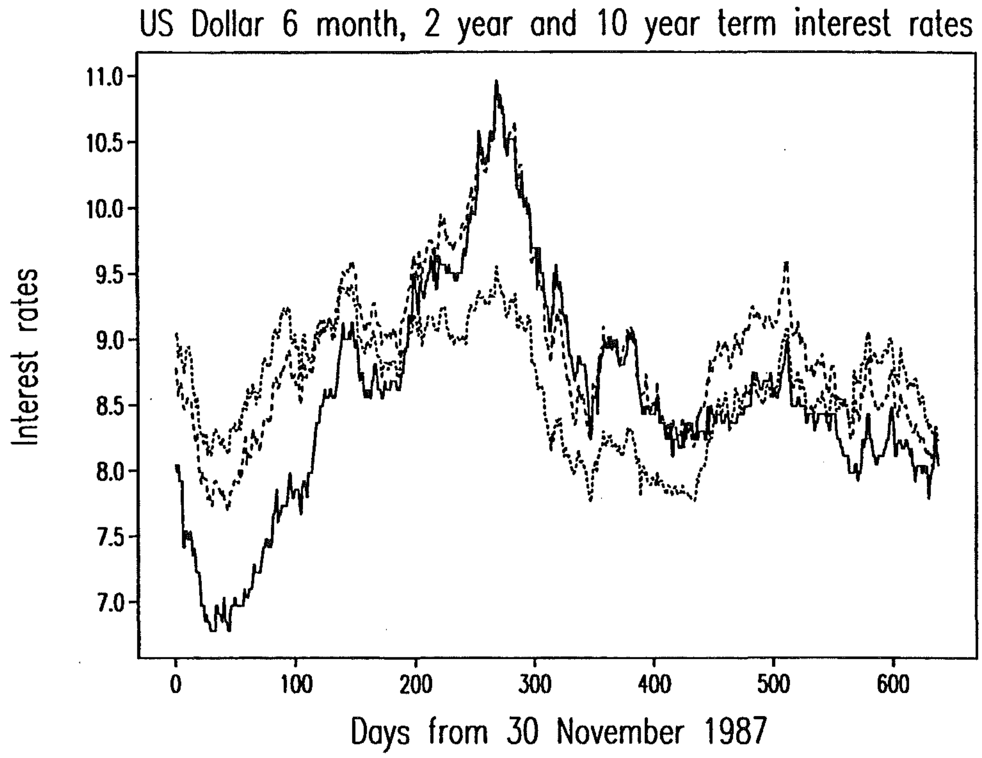


Figure 2: Six-month (solid line), two year (broken line) and ten year (dotted line) dollar term rate series.

we need not be concerned to use non-standard distributions at the stage of identifying conditional independence graphs for the series.

In the next section we shall review the second and third stage of our procedure; how we form and use the sample partial autocorrelation coefficients of the series to construct (tentative) conditional independence graphs, and how we use these to identify possible structural VAR models for stationary processes.

In section 3 we consider how the definition of conditional independence graphs and sample partial correlation coefficients can be usefully extended to $I(1)$ processes, and how they are to be interpreted when theoretical autocorrelations and partial autocorrelations are not defined. We also extend to the $I(1)$ case, the moralization rule that allows us to determine which DAG representations, of structural VAR models, are consistent with the CIG. In section 4 we present a theorem which shows that the statistical methods for constructing a CIG for the series in the stationary case, may also be applied in the case of $I(1)$ vector autoregressions. Our final section will present a simulation of a cointegrated $I(1)$ vector AR(1) process with 7 components, that illustrates the validity of this result.

# 2 Use of the conditional independence graph for model selection in the stationary case.

## 2.1 Construction of the conditional independence graph.

The statistical methods are based on a data matrix $X$ which in the general case consists of $m(P+1)$ vectors of length $n = N-P$, composed of elements $x_{t-u,i}$, $t = P+1-u, \dots N-u$, for each series $i = 1, 2, \dots, m$, and each lag $u = 0, 1, \dots, P$, for some chosen maximum lag $P$. Our first stage is that of overall order selection. For each order $p$ we fit, by OLS, the saturated structural VAR regressions of the $m$ contemporaneous (lag 0) vectors on all the vectors up to lag $p$. Using the sums of squares $S_i$ from these regressions we form the AIC as $n \sum \log S_i + 2k$, where $k = pm^2 + m(m-1)/2$ is the total number of regression coefficients estimated in the regressions. For the saturated model the causal order of the contemporaneous variables does not affect the result. Each one is included only as a regression variable for a subsequent variable in the chosen ordering. We then select the order $p$ which minimizes the AIC. In this way we selected the order $p = 2$ for the Italian monetary series, which is in agreement with that found in a previous analysis by Bagliano and Favero (1998).

Our second stage is to construct the sample CIG for the chosen model order $p$. This CIG consists of the same nodes as those shown in Figure 1, representing the variables up to lag 2. In general a CIG is an undirected graph, defined by the *absence* of a link between two nodes if they are independent conditional upon *all* the remaining variables. Otherwise the nodes are linked. In a Gaussian context this conditional independence is indicated by a zero partial autocorrelation:

$$\rho\left(x_{t-u,i}, x_{t-v,j} \middle| \{x_{t-w,k}\}\right) = 0, \tag{4}$$

where the set of conditioning variables on the right is the whole set up to lag $p$, excluding the variables on the left. As shown by Whittaker (1990), the set of all such partial correlations required to construct the CIG is conveniently calculated from the inverse $W$, of the covariance matrix $V$ of the whole set of variables, as

$$\rho\left(x_{t-u,i}, x_{t-v,j} \middle| \{x_{t-w,k}\}\right) = -W_{hl} / \sqrt{(W_{hh} W_{ll})} \tag{5}$$

where $h$ and $l$ respectively index the lagged variables $x_{t-u,i}$ and $x_{t-v,j}$ in the matrices $V$ and $W$. In the wider linear least squares context, defining linear partial autocorrelations as the same function of linear unconditional correlations as in the Gaussian context, (4) still usefully indicates lack of linear predictability of one variable by the other given the inclusion of all remaining variables.

## 2.2 Sample partial autocorrelations.

To obtain estimates $\hat{\rho}$ of the partial autocorrelations, we use in place of $V$ the sample covariance matrix $\hat{V}$ formed from the data matrix $X$, but including only lags up to $p$.

We then need a statistical test to decide which links are absent in the CIG. We are only concerned with links between contemporaneous variables and between contemporaneous and lagged variables, because these are the only ones that appear in the structural model DAG. The test we use is to retain a link when the sample partial autocorrelation $\hat{\rho}$ satisfies:

$$|\hat{\rho}| > z/\sqrt{(z^2 + \nu)}), \tag{6}$$

where $z$ is an appropriate critical value of the standard normal distribution and $\nu =$ is a residual degrees of freedom in the regression of any one column of the data matrix $X$ on all the other columns.

This derives from two results. The first is the algebraic relationship relating a sample partial correlation $\hat{\rho}$ between any two regression vectors of the data matrix $X$, to the $t$ value of one such vector in the regression of the other vector on that, and all the other vectors of $X$. This is given by $\hat{\rho} = t/\sqrt{(t^2 + \nu)}$ (see Greene, 1993,p 180).

The second result is the asymptotic normal distribution of the $t$ value for time series regression coefficients, given for example by Anderson (1971, p211). Of course we should properly apply multiple testing procedures when applying the test simultaneously to all sample partial autocorrelations, but that is not a practical option. Our attitude is similar to that advocated by Box and Jenkins (1976) for the identification, for example, of autoregressive models using time series partial autocorrelations. We use these values to suggest possible models; after fitting these we apply more formal tests and diagnostic checks to converge on an acceptable model.

## 2.3 Selection of a structural VAR model.

To return to our example, the critical value for significance at the 5% level is 0.207. Figure 3 shows the appropriate subgraph of the CIG of the lagged variables constructed using this threshold, with the addition of two links, $x_{t,3} - x_{t-1,1}$ and $x_{t,3} - x_{t-2,2}$ shown by broken lines. These are included because their partial autocorrelations are very close to the threshold. Our third stage is to determine which DAG representations are consistent with the CIG in (3), or are nearly so, allowing for statistical uncertainty. For this purpose we use, in an inverse manner, the *moralization* rule of Lauritzen and Spiegelhalter (1988), by which we can form the CIG that would arise from any hypothesized DAG interpretation. This rule, is to insert an undirected link between any two nodes $a$ and $b$ for which there is a node $c$ with directed links both $a \rightarrow c$ and $b \rightarrow c$. In this case $c$ is known as a common child of $a$ and $b$, and the insertion of a new, moral, link as *marrying* $a$ and $b$, which are known as the parents of $c$. After doing this for the whole graph the directions are removed from the original links. The original links therefore remain, and the only new links are the moral links.

Of course we attach the arrow of time to links from the past to the present, so the challenge is to clarify the directions of the recursive ordering of contemporaneous variables. As we describe in Reale and Tunnicliffe Wilson, (2001), inspection of Figure 3 lead us swiftly to the specification of a good structural model for these series, as represented in Figure 1.

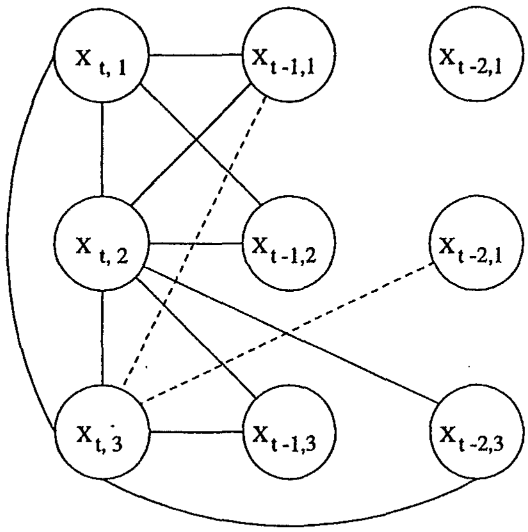


Figure 3: The CIG estimated for the Italian monetary series.

Moralization of this graph gives a very close approximation to Figure 3. On the evidence of the AIC, the model in Figure 1 represents the structure as well as the full VAR(2) model with fewer than half the parameters, and with improved predictive ability.

# 3 The graphical representations of $I(1)$ processes.

## 3.1 Definitions of the partial DAG and CIG.

We now assume that our series follows a vector $I(1)$ process with an arbitrary degree of cointegration, that is represented by some *true* structural VAR($p$) model of the form (1). As in the stationary case we can transform this to the canonical VAR($p$) form (2), and it is to this representation that we would apply the usual conditions for cointegration, as set out, for example, by Reinsel, (1993), p 164.

For this process we now wish to define partial graphs on the nodes that are the components of $x_t$, $x_{t-1}$, ..., $x_{t-p}$, i.e. we are concerned only with the presence or absence of links between elements of $x_t$ and from elements of $x_t$ to elements of lagged terms, $x_{t-h}$. For the purpose of identifying a structural VAR representation of the process, we are *not* here concerned with links between lagged terms.

The representation of (1) by a partial DAG, proceeds exactly as for the stationary case, because the process is fully determined by the specified structural VAR model. However, we have difficulty in defining the CIG when the process is $I(1)$, because there is no fixed distribution that we can assume for the variables that constitute the nodes. The VAR model specifies only the conditional distribution of $x_t$ given $x_{t-1}$, ..., $x_{t-p}$. We might proceed, for any given time $t$, by assuming that $x_{t-1}$, ..., $x_{t-p}$ have a well defined joint Gaussian distribution, though this would be unknown, and changing with time. We can avoid this assumption by basing all our procedures on the conditional distribution specified by the model. For the assumed cointegrated model, expressed in the canonical form (2), we therefore define the links of the required partial CIG as follows.

**Definition 1.** *The partial CIG for a structural VAR model.*

A link is *absent* between $x_{t,i}$ and $x_{t-h,j}$, where $0 \leq h \leq p$, if and only if the coefficient of $x_{t-h,j}$ is *zero*, in the regression of $x_{t,i}$ upon all the elements of $x_t$, $x_{t-1}$, ..., $x_{t-p}$ (other than $x_{t,i}$ itself).

*Remark* We are considering here *not* the regression that might be estimated from the data (which comes later), but the regression that may be determined from any given structural or canonical cointegrated VAR($p$) model that might be considered to represent the process. This regression may be determined from the canonical form by (i) choosing any re-ordering of the elements of $x_t$ in which $x_{t,i}$ comes first, (ii) determining the matrix $\Phi_0$ that corresponds to this re-ordering and satisfies (3), and (iii) forming the corresponding structural model (1) by premultiplying (2) by $\Phi_0$. The required coefficient is then $\Phi_{h,i,j}$ in (2). Note that this structural representation would *not* in general be the *true* sparse form that our procedure is designed to identify. This definition extends to I(1) processes, the definition for a stationary VAR in terms of the partial autocorrelation, that the link is absent if and only if

$$\rho\left(x_{t,i}, x_{t-h,j} | \{x_{t-\ell,k}\}\right) = 0, \tag{7}$$

where the conditioning set of variables on the right, includes all the elements of $x_t$, $x_{t-1}$,..., $x_{t-p}$, excluding the pair on the left. Both definitions imply that $x_{t-h,j}$ has no extra (linear) predictive capability for $x_{t,i}$, in addition to that of the conditioning variables.

## 3.2 The moralization rule.

The derivation of this rule in the general context of graphical models assumes that the joint distribution of the variables is well specified. Again, because the model (1) only specifies a limited set of conditional distributions, we cannot invoke it directly, but will prove that it still applies in this specific context, in conjunction with our extended the definition 1 of the partial CIG.

**Theorem 1**

Given the DAG corresponding to some assumed *true* cointegrated structural VAR(1), then the CIG determined for this model, using the definition 1 of the CIG given above, is exactly that formed by applying the moralization rule to the DAG.

*Remark* The moralization rule is a general one in the sense that it is possible in exceptional cases, through coincidence of numerical cancellation, that a link defined by the rule may in fact be absent. Such an example may be contrived by starting with an appropriate CIG. In the following proof we discount such cases. This is manifest as an assumption that a term in a matrix product will be non-zero if, in the calculation of that term, one of the contributing products is known to be non-zero.

**Proof.**

We start with a DAG representation of the structural model 1, for which we assume, without loss of generality, a contemporaneous causal structure that follows our standard ordering with $\Phi_0$ being upper triangular as shown in (8). In DAG terminology this means that any contemporaneous variable $x_{t,j}$ may only be a child of, i.e. dependent upon, either another contemporaneous variable $x_{t,i}$ for which $i > j$, or any lagged variable $x_{t-h,j}$. We derive the moralization rule by determining the children of $x_{i,t}$, when the contemporaneous causal structure is reordered, so that $x_{i,t}$ becomes the *first* variable, that is regressed upon *all* remaining contemporaneous variables and all lagged variables. To do this we derive from (1) an alternative structural VAR representation with coefficients $\tilde{\Phi}_0$, $\tilde{\Phi}_1, \ldots, \tilde{\Phi}_p$. In this representation the contemporaneous autoregressive matrix $\tilde{\Phi}_0$, also shown in (8), corresponds to a reversal of the ordering of $x_{t,1}$, $x_{t,2}$, ..., $x_{t,i}$. The top left partitions in (8) corresponds to these variables, with $T$ being upper triangular and $L$ lower triangular. The partitions $A$ and $B$ may be of saturated or sparse structure. All diagonal elements are unity.

$$\Phi_0 = \begin{pmatrix} \ddots & T & \vdots & & \\ 0 & \ddots & \vdots & V & \\ 0 & 0 & 0 & \ddots & \\ 0 & 0 & 0 & 0 & \ddots \end{pmatrix} ; \; \tilde{\Phi}_0 = \begin{pmatrix} \ddots & 0 & \vdots & & \\ L & \ddots & \vdots & W & \\ 0 & 0 & 0 & \ddots & \\ 0 & 0 & 0 & 0 & \ddots \end{pmatrix} . \tag{8}$$

Our result requires the following Lemma which we prove in the appendix.

**Lemma 1.** For all lags $h = 0, 1, \ldots p$,

$$\tilde{\Phi}_{hij} = \frac{\tilde{D}_{ii}}{D_{ii}} \Phi_{hij} + \sum_{s=1}^{i-1} \Phi_{0si} \frac{\tilde{D}_{ii}}{D_{ss}} \Phi_{hsj} \tag{9}$$

where $\tilde{D}$ is the diagonal matrix of the white noise process in the alternative model.

From this lemma we conclude that $\tilde{\Phi}_{hij} \neq 0$, i.e. there is a link from $x_{t-h,j}$ to $x_{t,i}$, if and only if either or both of the following hold:

- $\Phi_{hij} \neq 0$, i.e. $x_{t,i}$ was a child of $x_{t-h,j}$ in the original model,
- there exists an $s$ such that $\Phi_{0si} \neq 0$ and $\Phi_{hsj} \neq 0$, i.e. $x_{t,i}$ and $x_{t-h,j}$ have a common child $x_{t,s}$ in the original model. This completes our proof.

# 4 The use of partial autocorrelations for $I(1)$ processes.

Because the theoretical partial autocorrelations are not directly defined for the integrated process, we were lead to define the CIG for these processes in terms of regression coefficients.

Corresponding to this definition, in the sample context we now consider the standard OLS regression of the vector corresponding to $x_{t,i}$ in $X$, upon all other vectors of $X$, and the usual $t$ value formed for the coefficient of the vector corresponding to $x_{t-h,j}$. We include a link between $x_{t,i}$ and $x_{t-h,j}$ in the CIG only if this $t$ value is statistically significant, i.e. if $|t| > c$ for some suitable critical value $c$ under the null hypothesis that the regression coefficient is zero. The theorem that follows proves that asymptotically, for a cointegrated VAR($p$) process, this critical value is that of a standard normal variable, because we are able to apply the inference procedures for OLS.

To implement this test we shall, however, use the algebraically equivalent inequality for the sample autocorrelation coefficient expressed in (6). We shall continue to call these sample autocorrelations, even though the theoretical partial autocorrelations are not defined. They are formed from the sample covariance matrix $\hat{V}$ in exactly the same way as for the stationary case. The whole procedure for constructing the CIG is then identical to that presented for the stationary process.

**Theorem 2.** Let $x_t$ follow a cointegrated structural VAR($p$) process of the form given by (1), and $X$ be the data matrix formed of lagged series values as described in section 2. Perform the standard OLS regression of the vector corresponding to $x_{t,1}$ in $X$, upon all other vectors of $X$ and form the usual $t$ value for the estimate of the coefficient $\Phi_{h,1,j}$ of the vector corresponding to $x_{t-h,j}$. Then under the hypothesis that $\Phi_{h,1,j} = 0$, $t$ has an asymptotic standard normal distribution.

**Remark.**For convenience of derivation and without loss of generality we take the response vector corresponding to $x_{t,1}$. Our development is based on results presented by Ahn and Reinsel (1990) that are set out also in Reinsel (1993), pp 165-8. We follow this with modifications (and some notational changes) to incorporate the structural form of model.

**Proof.**

(i) Write the model (1) in error correction form using the differenced series $w_t = x_t - x_{t-1}$, as

$$w_t = Cx_{t-1} + \varphi_0 w_t + \sum_{j=1}^{p-1} \varphi_j w_{t-j} + a_t, \tag{10}$$

so that the original parameters may be expressed in terms of the new ones as

$$\begin{aligned} \Phi_0 &= I - \varphi_0 \\ \Phi_1 &= \varphi_1 + I - \varphi_0 + C \\ \Phi_k &= \varphi_k - \varphi_{k-1} \quad k = 2, \ldots, p-1, \\ \Phi_p &= -\varphi_{p-1}. \end{aligned} \tag{11}$$

We require $\varphi_0$ to be zero on and below the diagonal, but $C$ and $\varphi_1, \ldots, \varphi_{p-1}$ are not restricted, except by cointegrating conditions to be specified. We shall also use the expression for

$$C = -(\Phi_0 - \sum_{k=1}^{p} \Phi_k). \tag{12}$$

(ii) The model may be transformed to the canonical form (2) in which the cointegratin condition may be expressed by two statements, Firstly that

$$\det\left(I - \sum_{k=1}^{p} \Phi_k^* B^k\right) = 0 \tag{13}$$

has $d < m$ roots equal to unity with all other $r = m - d$ roots outside the unit circl Secondly that there exist matrices $P$ and $Q = P^{-1}$ such that

$$\sum_{k=1}^{p} \Phi_k^* = P \begin{pmatrix} I_d & 0 \\ 0 & \Lambda_r \end{pmatrix} Q, \tag{14}$$

where $\Lambda_r$ is of Jordan canonical form with diagonal elements less than one in absolut value. We can then express

$$x_t = P\, z_t = \begin{pmatrix} P_1 & P_2 \end{pmatrix} \begin{pmatrix} z_{1,t} \\ z_{2,t} \end{pmatrix} = P_1 z_{1,t} + P_2 z_{2,t}, \tag{15}$$

where $z_{1,t}$ is a $d$ dimensional purely integrated process and $z_{2,t}$ is an $r$ dimensiona stationary series, defined by

$$z_t = \begin{pmatrix} z_{1,t} \\ z_{2,t} \end{pmatrix} = Q\, x_t = \begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix} x_t. \tag{16}$$

Using (15) we can express the error correction form (10) as

$$w_t = C_1\, z_{1,t-1} + C_2\, z_{2,t-1} + \varphi_0\, w_t + \sum_{j=1}^{p-1} \varphi_j\, w_{t-j} + a_t, \tag{17}$$

where $C_1 = C\, P_1$ and $C_2 = C\, P_2$. Conversely, $C = C_1\, Q_1 + C_2\, Q_2$.

With the second equation in (11) replaced by

$$\Phi_1 \;=\; \varphi_1 + I - \varphi_0 + C_1\, Q_1 + C_2\, Q_2 \tag{18}$$

we now have a 1:1 linear transformation of parameters from $C_1$, $C_2$, $\varphi_0, \ldots, \varphi_{p-1}$, t $\Phi_0, \ldots, \Phi_p$. Although the transformation coefficients $Q_1$ and $Q_2$ in (18) are mode dependent, we know sufficient to establish our result.

(iii) The true value of $C_1$ is zero. From (14)

$$\begin{aligned} (I - \textstyle\sum_{k=1}^{p} \Phi_k^*) \begin{pmatrix} P_1 & P_2 \end{pmatrix} &= \begin{pmatrix} P_1 & P_2 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & I_r - \Lambda_r \end{pmatrix} \\ &= \begin{pmatrix} 0 & P_2(I_r - \Lambda_r) \end{pmatrix}, \end{aligned} \tag{19}$$

so that

$$(I - \sum_{k=1}^{p} \Phi_k^*)P_1 = 0. \tag{20}$$

Then from (12)

$$\begin{aligned} C P_1 &= -(\Phi_0 - \textstyle\sum_{k=1}^{p} \Phi_k)P_1 \\ &= -\Phi_0 (I - \textstyle\sum_{k=1}^{p} \Phi_k^*)P_1 \\ &= 0. \end{aligned} \tag{21}$$

(iv) By the arguments presented in Reinsel (1993) we can establish the properties of the estimated coefficients in (17), based on the fact that the regressors $z_{2,t}$ are purely $I(1)$ and the remainder are stationary. Because the true value of $C_1$ is zero, we have

$$\hat{C}_1 = O(N^{-1}) \tag{22}$$

and

$$\left(\hat{C}_2 - C_2,\ \hat{\varphi}_0 - \varphi_0,\ \ldots,\ \hat{\varphi}_{p-1} - \varphi_{p-1}\right) = O(N^{\frac{1}{2}}). \tag{23}$$

Furthermore, the joint distribution of the estimated coefficients in (23) is asymptotically the same as for a purely stationary model, i.e. it is valid to use for these the inference procedures of OLS.

(v) Now consider the estimates of the original coefficients $\Phi_j$ listed in (11). Apart from $\Phi_1$ we see that the estimates are linear transformations of those in (23), so we may also apply to these the inference procedures of OLS. For $\hat{\Phi}_1$ we refer to (18). The magnitude of $\hat{C}_1$ is $O(N^{-1})$, so provided that the particular element of $\hat{\Phi}_1$ in which we are interested also contains components from $\hat{\varphi}_0$, $\hat{\varphi}_1$ or $\hat{C}_2$, these, having standard error of magnitude $O(N^{-\frac{1}{2}})$ will dominate the distribution of the estimate for large $N$, when OLS based inference may again be used.

The only case when the first two of these components are not present arises when $p = 1$, so that $\varphi_1$ is not present, and when the parameter of interest is $\Phi_{1,1,1}$, because there is no component $\Phi_{0,1,1}$ in the structural model. We complete the proof by showing that even in this final case, $\hat{\Phi}_{1,1,1}$ will always contain a component from $\hat{C}_2$, under our null hypothesis, that becomes $\Phi_{1,1,1} = 0$.

The component of $\hat{\Phi}_{1,1,1}$ contributed by $\hat{C}_2$ is, from (18), $\hat{c}_2 q_2$, where $\hat{c}_2$ is the first row of $\hat{C}_2$ and $q_2$ is the first column of $Q_2$. We need only show that $q_2 \neq 0$ to verify that $\hat{c}_2 q_2$ will contribute a component of magnitude $O(N^{-\frac{1}{2}})$ to $\hat{\Phi}_{1,1,1}$. Now from (14), and using that $p = 1$, we find

$$(I - \Phi_1^*)_{\text{column }1} = P_2(I_r - \Lambda_r)q_2 \tag{24}$$

Thus $q_2 = 0$ implies that

$$(\Phi_1^*)_{\text{column }1} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{25}$$

But because $\Phi_1 = \Phi_0 \Phi_1^*$, and $\Phi_0$ has unit diagonals, this implies that $\Phi_{1,1,1} = 0$, so contradicting the null hypothesis in this final case, and demonstrating that in fact $q_2$ cannot be a zero column. This completes our proof.

We conclude this section by noting that the stationary regressors in (17) are either differences of $x_t$ or stationary components of $x_t$, none of which we would expect to be *near-*$I(1)$. We would therefore expect that acceptable normality of the OLS $t$ statistics would be achieved for moderate series length $N$

# 5 An illustrative simulation example.

To illustrate the previous result we simulated a cointegrated sparse structural VAR(1) model of seven series:

$$\Phi_0 x_t = \Phi_1 x_{t-1} + a_t \tag{26}$$

with standard deviations of the components of $a_t$ being, in order, 0.0457, 0.0646, 0.0782, 0.0197, 0.0336, 0.0178, and 0.0164.

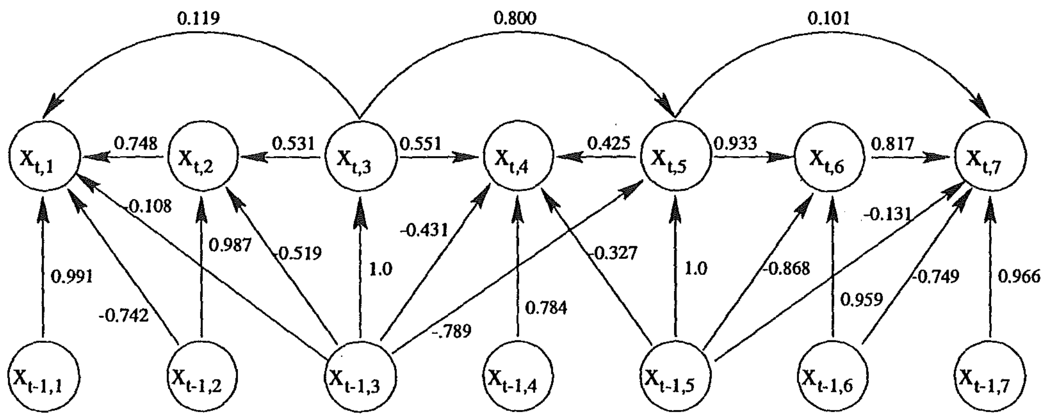


Figure 4: The DAG representation of the simulated structural VAR(7) model.

This model is represented by the DAG in Figure 4, where the numbers associated with the arrows are the coefficients of the non-zero autoregressive coefficients. We derived this model from the AR part of the stationary structural VARMA(1,1) that we fitted to the seven term rate series in Tunnicliffe Wilson, Reale, and Morton, (2001). It is capable of producing simulated series that are very similar to those in Figure 2.

However, we have slightly adjusted, to the value of unity, the coefficients that relate $x_{t,3}$ and $x_{t,5}$ to their respective lagged values $x_{t-1,3}$ and $x_{t-1,5}$. Each series $x_{t,i}$ is therefore individually $I(1)$, but the vector series has 5 cointegrating vectors, i.e. there are 2 $I(1)$ components and 5 stationary components of $x_t$.

The CIG derived by moralization of Figure 4 is shown in Figure 5. We selected ten pairs of variables that are not linked in Figure 5 and computed the sample partial autocorrelation coefficient between these pairs for 10,000 replications of simulated series of length 600. The numbers of values exceeding the 5% critical threshold defined by (6) is shown in table 1.

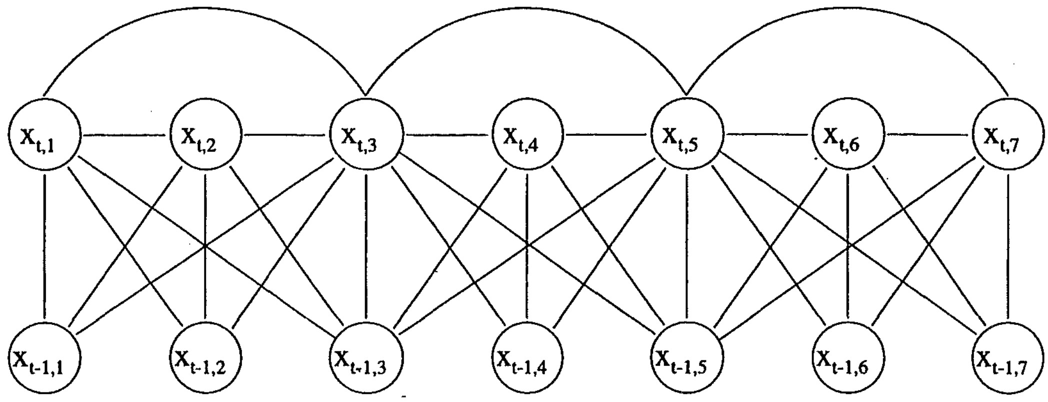


Figure 5: The CIG derived from the DAG representation of the simulated structural VAR(7) model.

These lie generally within the expected range of variability about the value of 500 that is expected asymptotically, and provides substantial support for the theory developed in the previous section.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Variable 1 | $x_{t,2}$ | $x_{t,3}$ | $x_{t,4}$ | $x_{t,2}$ | $x_{t,3}$ | $x_{t,4}$ | $x_{t,4}$ | $x_{t,4}$ | $x_{t,5}$ | $x_{t,6}$ |
| Variable 2 | $x_{t,4}$ | $x_{t,6}$ | $x_{t,6}$ | $x_{t-1,4}$ | $x_{t-1,6}$ | $x_{t-1,1}$ | $x_{t-1,2}$ | $x_{t-1,6}$ | $x_{t-1,2}$ | $x_{t-1,3}$ |
| Numbers | 440 | 556 | 510 | 497 | 550 | 502 | 464 | 529 | 485 | 548 |

Table 1: The numbers of selected sample partial autocorrelations that exceed critical thresholds in 10,000 simulations.

# Appendix

**Proof of Lemma 1.** We may write, for $h = 0, 1, \ldots p$,

$$\tilde{\Phi}_h = \begin{pmatrix} R & \vdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \vdots & I \end{pmatrix} \Phi_h \tag{27}$$

where $R = LT^{-1}$, because, for $h = 0$, this transforms $\Phi_0$ to the form of $\tilde{\Phi}_0$ shown in (8), and that form is unique. Let $P$ be the upper left $i \times i$ partition of $\Sigma^{-1}$, where $\Sigma$ is the variance matrix of the linear innovation in the canonical model (2). From (3),

$$T'ET = L'\tilde{E}L = P, \tag{28}$$

where $E$ and $\tilde{E}$ are the inverses of the corresponding upper left partitions of $D^{-1}$ and $\tilde{D}^{-1}$. Note that, because $L'$ is *upper* triangular, with unit diagonals, the last, $i$th, row of $P$ is

$$P_{\text{row } i} = (L')_{\text{row } i}\tilde{E}L = \tilde{E}_{i,i}L_{\text{row i}}. \tag{29}$$

Now consider the last row of $R$:

$$R_{\text{row } i} = L_{\text{row } i} T^{-1} = \tilde{E}_{i,i}^{-1} P_{\text{row } i} T^{-1} = \tilde{E}_{i,i}^{-1} \left( P T^{-1} \right)_{\text{row } i} = \tilde{E}_{i,i}^{-1} \left( T' E \right)_{\text{row} i}, \qquad (30)$$

which has elements $\Phi_{0\,si} \frac{\tilde{D}_{ii}}{D_{ss}}$, the last, for $s = i$, reducing to $\frac{\tilde{D}_{ii}}{D_{ii}}$. The result of the Lemma comes from using this last expression in row $i$ of (27).

# References


Abadir, K. M., 1995, The limiting distribution of the $t$ ratio under a unit root, Econometric Theory, 11, 775-793.

Ahn, S. K., and Reinsel, G. C. 1990. Estimation for partially nonstationary multivariate autoregressive models, Journal of the American Statistical Association 85, 813–823.

Akaike, H., 1973, A new look at Statistical Model Identification, IEEE Transactions on Automatic Control AC-19, 716–723.

Anderson, T.W., 1971, The statistical analysis of time series (Wiley, New York).

Bagliano, F.C. and A.C. Favero, 1998, Il canale del credito della politica monetaria. Il caso Italia, in: G. Vaciago, ed., Moneta e Finanza (Il Mulino, Bologna).

Box, G.E.P. and G.M. Jenkins, 1976, Time series analysis, forecasting and control (Holden-Day, Oakland).

Dahlhaus, R., 2000, Graphical interaction models for multivariate time series, Metrika 51, 157–172.

Glymour, C. and P. Spirtes, 1988, Latent variables, causal models and overidentifying constraints, Journal of Econometrics 39, 175–198.

Greene, W.H., 1993, Econometric analysis (Prentice-Hall, Englewood Cliffs).

Lauritzen, L.S and T.S. Richardson, 2002, Chain graph models and their causal interpretation, Journal of the Royal Statistical Society B 64, 321–361.

Lauritzen, S.L. and D.J. Spiegelhalter, 1988, Local computations with probabilities on graphical structures and their applications to expert systems, Journal of the Royal Statistical Society B 50, 157–224.

Pearl, J., 2000, Causality (Cambridge University Press, Cambridge, UK).

Reale, M. and G. Tunnicliffe Wilson, 2001, Identification of vector AR models with recursive structural errors using conditional independence graphs, Statistical Methods and Applications 10, 49–65.

Reale, M. and G. Tunnicliffe Wilson, 2002, The sampling properties of conditional independence graphs for structural vector autoregressions, Biometrika 89, 457–461.

Reinsel, G. C., 1993, Elements of Multivariate Time Series Analysis, (Springer-Verlag, New York)

Spirtes, P., C. Glymour and R. Scheines, 2000, Causation, prediction and search (MIT Press, Cambridge, MA).

Swanson, N.R. and C.W.J. Granger, 1997, Impulse response functions based on a causal approach to residual orthogonalization in vector autoregressions, Journal of the American Statistical Association 92, 357–367.

Tunnicliffe Wilson, G., M. Reale and A.S. Morton, 2001, Developments in Multivariate Time Series Modeling, Estadistica 53, 353–395.

Whittaker, J.C., 1990, Graphical Models in Applied Multivariate Statistics (Wiley, Chichester).